






## Sample generation for the spin-fermion model using neural networks

Georgios Stratis <sup>1,2,\*</sup>, Phillip Weinberg,<sup>3</sup> Tales Imbiriba <sup>1</sup>, Pau Closas <sup>1</sup> and Adrian E. Feiguin <sup>3</sup>

<sup>1</sup>*Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115, USA*

<sup>2</sup>*Roux Institute, Northeastern University, Portland, Maine 04101, USA*

<sup>3</sup>*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*

 (Received 15 June 2022; revised 21 October 2022; accepted 31 October 2022; published 8 November 2022)

Monte Carlo simulations of hybrid quantum-classical models such as the double exchange Hamiltonian require calculating the density of states of the quantum degrees of freedom at every step. Unfortunately, the computational complexity of exact diagonalization grows as a function of the system's size  $N$ , making it prohibitively expensive for any realistic system. We consider leveraging data-driven methods, namely, neural networks, to replace the exact diagonalization step in order to speed up sample generation. We explore a model that learns the free energy for each spin configuration and a second one that learns the Hamiltonian's eigenvalues. We implement data augmentation by taking advantage of the Hamiltonian's symmetries to artificially enlarge our training set and benchmark the different models by evaluating several thermodynamic quantities. While all models considered here perform exceedingly well in the one-dimensional case, only the neural network that outputs the eigenvalues is able to capture the right behavior in two dimensions. The simplicity of the architecture we use in conjunction with the model agnostic form of the neural networks can enable fast sample generation without the need of a researcher's intervention.

DOI: [10.1103/PhysRevB.106.205112](https://doi.org/10.1103/PhysRevB.106.205112)

### I. INTRODUCTION

Strongly correlated materials are characterized by dominant electron-electron interactions [1–3] with electronic and magnetic properties that can potentially be tuned and used for novel technological applications beyond the silicon/semiconductor paradigm. To understand the properties of these systems, it is unavoidable to resort to numerical methods such as quantum Monte Carlo (QMC) [4,5]. In models unaffected by the infamous sign problem, QMC can treat very large systems with dozens, if not hundreds, of degrees of freedom. One such scenario is the case of the double-exchange Hamiltonian [6–9] (also dubbed, in a different context, as the spin-fermion model) and multiorbital generalizations. These models have shed light on the physics of manganites with colossal magnetoresistance [10–12] and high-temperature superconductivity [13]. In such hybrid quantum-classical Hamiltonians, conduction electrons are coupled to classical degrees of freedom, spins in this case. During the Monte Carlo sample generation for the spin-fermion model, one needs to calculate the eigenvalues of the electronic Hamiltonian, a task that can be done exactly using matrix diagonalization. Unfortunately, the computational complexity of exact diagonalization scales unfavorably for a matrix of size  $N$ , limiting the simulations to a couple of hundred orbitals. Furthermore, the number of updates needed to obtain uncorrelated Monte Carlo samples can be large close to a phase transition. Thus in order to study larger systems or systems close to a phase transition we need to find faster

alternatives to exact diagonalization. There have been various attempts to overcome this issue, such as expanding the density of states using Chebyshev polynomials [14–17], obtaining the Hamiltonian's eigenspectrum corrections after a local update [18], and combining the kernel polynomial method with Langevin dynamics [19]. An alternative approach, which is explored in this paper, is to use data-driven machine learning methods to assist with this computational task.

In the last five years, machine learning has found many applications in quantum physics, such as learning thermodynamics [20], finding the ground state of a many-body system [21], identifying phase transitions [22], conducting quantum state tomography [23], and calculating spectral functions [24,25]. Moreover, approaches to accelerate Monte Carlo simulations with machine learning have been explored in the self-learning Monte Carlo method [26], using restricted Boltzmann machines [27], deep [28], and autoregressive neural networks [29]. In this work, we design two neural network models, both featuring simple architectures, that generate samples to determine quantities of interest over a range of temperatures for the spin-fermion/double-exchange model. The first neural network outputs the free energy associated with the quantum degrees of freedom, whereas the second outputs the eigenvalues of the Hamiltonian for a given classical spin configuration. Both networks are trained using a data set that is artificially augmented by taking advantage of the system's translation and rotation symmetries. The manuscript is organized as follows: in Sec. II, we present the spin-fermion/double-exchange model, in Sec. III, we present the effective models and the training methodology used, in Secs. IV and V, we present our results, and we finally conclude with a discussion in Sec. VI.

\*g.stratis@northeastern.edu

## II. SPIN-FERMION MODEL

The spin-fermion model, also known as the double-exchange model, was first proposed by Zener [6] to explain ferromagnetism in materials with incomplete  $d$  shells and itinerant conduction electrons. It has been extensively studied [7–9] due to its relation with colossal magnetoresistance in manganites [10–12]. In the spin-fermion model, one treats each atom with its  $t_{2g}$  electrons as a classical localized spin  $\mathbf{S}_i \in \mathbb{R}^3$  at site  $i$  with unit magnitude  $|\mathbf{S}_i| = 1$ . The  $e_g$  conduction electrons are treated quantum mechanically and are able to move freely interacting only with the classical spin at the site they are located. The Hamiltonian for the spin-fermion model considered in this work is the following single-orbital version:

$$\hat{H}(\mathbf{S}) = -t \sum_{\langle ij \rangle, \sigma} (\hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + \hat{c}_{j\sigma}^\dagger \hat{c}_{i\sigma}) + \frac{J}{2} \sum_{i, \alpha, \beta, \gamma} S_i^\gamma \hat{c}_{i\alpha}^\dagger \sigma_{\alpha\beta}^\gamma \hat{c}_{i\beta}, \quad (1)$$

where  $\mathbf{S} = \{\mathbf{S}_1, \dots, \mathbf{S}_N\}$  is the classical spin configuration,  $t$  is the hopping constant and our unit of energy,  $\hat{c}_{i\sigma}^\dagger$  ( $\hat{c}_{i\sigma}$ ) is the fermionic creation (annihilation) operator at the  $i$ th site for fermion with spin  $\sigma \in \{\uparrow, \downarrow\}$ ,  $\langle ij \rangle$  are the pairs of nearest neighbors,  $J$  is the interaction strength between the classical spins and the electrons (Hund coupling),  $S_i^\gamma$  is the  $\gamma$ th component of the classical spin at the  $i$ th site, and  $\{\sigma^x, \sigma^y, \sigma^z\}$  are the Pauli matrices. The spin-fermion Hamiltonian has  $O(3)$  rotational symmetry in the spin sector and, because of the periodic boundary conditions we imposed, it also has discrete translation invariance. We use these two symmetries to improve the accuracy of the neural networks performance as described in Sec. III.

Due to the fact that our system contains both quantum and classical degrees of freedom, the partition function  $Z$

associated with the spin-fermion model is

$$Z = \int d\mathbf{S} \text{Tr}[\exp(-\beta(\hat{H} - \mu\hat{N}))] \\ = \int d\mathbf{S} \prod_{\nu} [1 + e^{-\beta(\epsilon_{\nu}(\mathbf{S}) - \mu)}], \quad (2)$$

where  $\beta = 1/T$  is the inverse temperature,  $\mu$  is the chemical potential,  $\hat{N}$  is the number operator, and  $\epsilon_{\nu}(\mathbf{S})$  are the single-particle eigenenergies corresponding to the Hamiltonian with spin configuration  $\mathbf{S}$ . The second equality follows from the fact that the electronic degrees of freedom are noninteracting and we are treating them with the grand-canonical ensemble. Given Eq. (2), one can define the probability of finding the system in a given classical spin configuration  $\mathbf{S}$  as

$$p(\beta, \mathbf{S}) = \frac{e^{-\beta\mathcal{F}(\beta, \mathbf{S})}}{Z}, \quad (3)$$

with  $\mathcal{F}(\beta, \mathbf{S})$  being the fermionic free energy given by

$$\mathcal{F}(\beta, \mathbf{S}) = -\frac{1}{\beta} \sum_{\nu} \ln[1 + e^{-\beta(\epsilon_{\nu}(\mathbf{S}) - \mu)}]. \quad (4)$$

Knowing the partition function enables us to determine the expected values for different quantities at various temperatures such as the average energy

$$\langle E \rangle = -\frac{\partial \ln(Z)}{\partial \beta} \\ = \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \sum_{\xi} \frac{\epsilon_{\xi}(\mathbf{S}) - \mu}{1 + e^{\beta(\epsilon_{\xi}(\mathbf{S}) - \mu)}} \\ = \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \sum_{\xi} [\epsilon_{\xi}(\mathbf{S}) - \mu] \rho(\beta, \epsilon_{\xi}, \mu), \quad (5)$$

and the specific heat

$$C_V = -\frac{\partial \langle E \rangle}{\partial T} = k\beta^2 \left\{ \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \left[ \left( \sum_{\xi} (\epsilon_{\xi}(\mathbf{S}) - \mu) \rho(\beta, \epsilon_{\xi}, \mu) \right)^2 \right. \right. \\ \left. \left. + \sum_{\xi} \left( (\epsilon_{\xi}(\mathbf{S}) - \mu) \right)^2 \rho(\beta, \epsilon_{\xi}, \mu) (1 - \rho(\beta, \epsilon_{\xi}, \mu)) \right] - \langle E \rangle^2 \right\}, \quad (6)$$

where  $\rho(\beta, \epsilon, \mu)$  is the Fermi-Dirac distribution. Furthermore, using Eqs. (2) and (3), we can calculate the magnitude of the average magnetization

$$|\mathbf{M}| = \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \left| \frac{1}{N} \left( \sum_i \hat{x} S_i^x + \hat{y} S_i^y + \hat{z} S_i^z \right) \right|, \quad (7)$$

and the staggered magnetization of our system

$$|\mathbf{M}_s| = \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \\ \times \left| \frac{1}{N} \sum_i (-1)^{\sum_{j=1}^d x_j} (\hat{x} S_i^x + \hat{y} S_i^y + \hat{z} S_i^z) \right|, \quad (8)$$

where  $x_j$  in the sum  $\sum_{j=1}^d x_j$  indicates the site index in the  $j$ th dimension. The parameters used in this paper favor antiferromagnetic order at low temperatures and this behavior should be captured by the staggered magnetization. Based on how we chose to define staggered magnetization, a  $|\mathbf{M}_s| = 1$  will indicate a situation where each site has spin with the opposite sign compared with its nearest neighbors. Lastly, it is customary to calculate the spin correlation

$$C(\mathbf{r}) = \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \frac{1}{N} \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+\mathbf{r}}, \quad (9)$$

and its Fourier transform, the spin structure factor

$$S(\mathbf{q}) = \frac{1}{Z} \int d\mathbf{S} e^{-\beta\mathcal{F}(\beta, \mathbf{S})} \sum_{\mathbf{r}} e^{-i\mathbf{q} \cdot \mathbf{r}} C(\mathbf{r}), \quad (10)$$

where  $\mathbf{q} = 2\pi(\frac{n_1}{L_1}, \dots, \frac{n_d}{L_d})$  with  $n_i = 0, 1, \dots, L_i - 1$  for a  $d$ -dimensional lattice  $L_1 \times L_2 \times \dots \times L_d$ . In all the thermal averages Eqs. (5)–(10), it is necessary to calculate the probability distribution given by Eq. (3). Unfortunately, one cannot find a closed form to Eq. (2), and has to resort to numerical methods such as Monte Carlo based approaches. The most straightforward way to do it is using exact diagonalization that is diagonalizing the Hamiltonian matrix to obtain the spectrum of eigenenergies associated with each spin configuration in order to determine its free energy. Once these are obtained, one can use the METROPOLIS Hastings algorithm [30] to generate samples that collectively approximate the real probability distribution, Eq. (3). The main drawback of exact diagonalization is the amount of resources needed scales unfavorably as the size of the system increases.

### III. MODEL TRAINING

In order to compare the different models, in this work we study the spin-fermion model in one-dimension using a system with  $N = 20$  lattice sites, and the two-dimensional counterpart using a system with  $N = 6 \times 6 = 36$  lattice sites. In both cases, we use interaction strength  $J = -1$  (ferromagnetic), and chemical potential  $\mu = 0$ . For each temperature, we generate training, validation, and testing data sets using three different METROPOLIS Hastings Markov chains to eliminate correlation between the data sets. Each data set contains spin configurations with their associated normalized free energy  $\frac{\mathcal{F}(\beta, \mathbf{S})}{N}$  and eigenvalues. For every Monte Carlo update, we use exact diagonalization to determine the free energy of the proposed spin configuration. Each Markov chain consists of a warm-up stage followed by a sample generation stage. During the warm-up stage we generate  $1000 \cdot N$  samples and determine the acceptance ratio  $r$ , which is the number of accepted over the total number of proposed spin configurations. The samples generated during the warm-up stage are not saved. In the sample generation stage, we generate  $N_s N / r$  samples and save only  $N_s$  of them. More specifically, for every  $N/r$  samples generated, we save only the last one. This is done to minimize correlations between samples. The training data set contains  $N_s = 10^4$  samples, whereas the validation and test data sets contained  $N_s = 10^3$  samples. We train three different models, an effective Heisenberg model and two neural network models. The effective Heisenberg model is used to benchmark the two neural networks against a well established approach based on physical insight.

#### A. Effective Heisenberg model

A simple model to approximate the free energy at a given temperature is a linear model inspired by Ruderman-Kittel-Kasuya-Yosida (RKKY) theory [31,32] where each classical spin interacts with every other classical spin via a Heisenberg interaction with a coupling strength that depends on the distance between the two spins [26,33]. The estimator  $\hat{\mathcal{F}}(\beta, \mathbf{S})$  of the free energy  $\mathcal{F}(\beta, \mathbf{S})$  is given by

$$\hat{\mathcal{F}}(\beta, \mathbf{S}) = \mathcal{F}_0 + \frac{1}{N} \sum_R J_R \sum_{\mathbf{r} \in \{|\mathbf{r}|=R\}} \sum_{\mathbf{i}} \mathbf{S}_{\mathbf{i}} \cdot \mathbf{S}_{\mathbf{i}+\mathbf{r}}, \quad (11)$$

where  $\mathcal{F}_0$  is a constant,  $N$  is the number of sites,  $J_R$  is the coupling strength between all classical spins that are separated by a distance  $R$  from the classical spin at site  $i$ . The constants  $\mathcal{F}_0$  and  $J_R$  are independent of the classical spin configuration, but depend on inverse temperature  $\beta$ . This model is linear due to the linear dependence between the free energy and the coupling strength constants  $J_R$  and arises from a perturbative treatment of the conduction electrons. We fit the model's parameters using linear regression for a least-squares problem whose solution is given by

$$\begin{pmatrix} \mathcal{F}_0 \\ \mathbf{J} \end{pmatrix} = (\tilde{\mathbf{C}}^\top \tilde{\mathbf{C}})^{-1} \tilde{\mathbf{C}}^\top \mathbf{F}, \quad (12)$$

where  $\mathbf{J}$  is the vector containing the coupling strength constants  $J_R$ ,  $\tilde{\mathbf{C}}_{m0} \equiv 1$ , the spin correlation corresponding to the distance  $|\mathbf{r}| = R$  for each sample

$$\tilde{\mathbf{C}}_{m,R} \equiv \frac{1}{N} \sum_{|\mathbf{r}|=R} \sum_{\mathbf{i}} \mathbf{S}_{\mathbf{i}}^{(m)} \cdot \mathbf{S}_{\mathbf{i}+\mathbf{r}}^{(m)}, \quad (13)$$

and  $\mathbf{F}_m$  stands for the free energy of training sample  $m$ . The columns of the matrix  $\tilde{\mathbf{C}}_{m,R}$  are organized in ascending order with the value  $R$  they represent. For instance, in the two-dimensional case, the first column corresponds to all the vectors with magnitude  $|\mathbf{r}| = 1$  ( $\mathbf{r} = \{(0, 1), (1, 0)\}$ ), and the second column corresponds to the vectors with magnitude  $|\mathbf{r}| = \sqrt{2}$  ( $\mathbf{r} = (1, 1)$ ). In essence, this effective Heisenberg model relates the spin correlation at different distances with the free energy, effectively tracing over the electronic degree of freedom and producing a much simpler approximation that is justified in the perturbative regime [31].

#### B. Neural network models

An artificial neural network is a function that maps its input from  $\mathbb{R}^m$  to its output in  $\mathbb{R}^n$ . In its simplest form a neural network consists of an input layer, a set of hidden layers, and an output layer where each layer consists of a set of nodes. The  $i^{\text{th}}$  node of the  $j^{\text{th}}$  hidden layer,  $h_i^{(j)}$ , takes as input the weighted sum of the outputs from the previous layer's nodes  $y_i^{(j)} = \sum_k w_{ik}^{(j)} z_k^{(j-1)}$ , and applies a nonlinear function  $f(y)$ , commonly referred to as the *activation function*, to generate its output  $z_i^{(j)} = f(y_i^{(j)})$ . In this work, without loss of generality, we use the Softplus activation function  $f(y) = \ln(1 + e^y)$ . To train a neural network for regression one needs to choose a loss function that compares the true value with the neural network's predicted value. We use the mean absolute error (MAE)  $\mathcal{L} = \frac{1}{N} \sum_i |y_i - \hat{y}_i|$  as our loss function and to find its minimum we use the Adam optimizer [34] with a learning rate that decays over time. At training epoch  $t$ , the learning rate is given by  $l_r(t) = l_r(0)\gamma^{-t}$ , where  $\gamma$  is the decay constant. We use initial learning rate  $l_r(0) = 10^{-3}$ , decay constant  $\gamma = 0.9995$ , and train for a total of  $N_e = 2 \times 10^4$  epochs. We noticed that in some cases neural networks with extremely low mean squared error generated samples that could accurately describe the system's average energy and specific heat, but the magnitude of the average magnetization departed significantly from the exact results. This issue is addressed by splitting the training set in minibatches during training, thus introducing some stochasticity in the optimization algorithm.

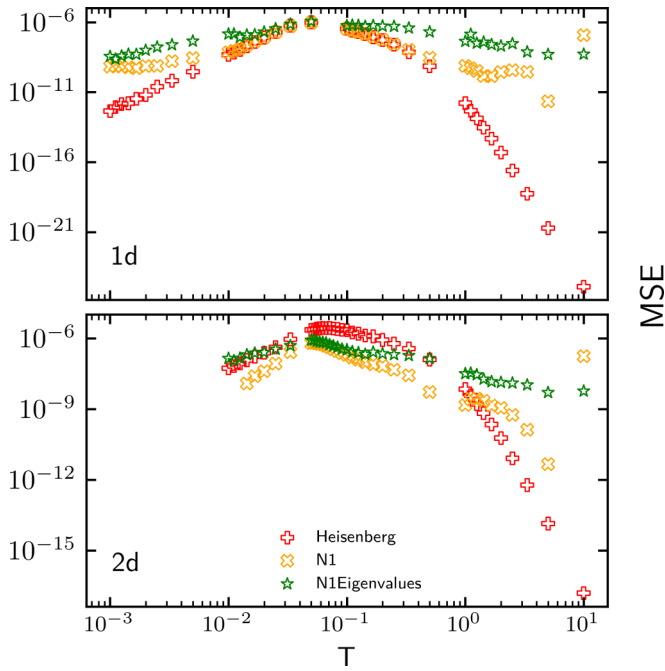


FIG. 1. Mean squared error between the exact and the predicted fermionic free energy for an one-dimensional lattice  $N = 20$  sites (top) and a two-dimensional lattice  $N = 6 \times 6 = 36$  sites (bottom).

In general, using a large number of minibatches improves a neural network's performance, but the training time grows as the number of minibatches increases. As a consequence, we split the training set in as many minibatches possible while making sure that the training time remains at a reasonable level.

The spin-fermion Hamiltonian has translational invariance, due to the periodic boundary conditions, and rotational invariance. We take advantage of these two symmetries to augment our training data, with the goal of improving the models' predictive performance. Data augmentation is a well established approach in machine learning due to its implementation simplicity [35]; however, the training time increases since the training set becomes larger. Through data augmentation one tries to teach a neural network the system's symmetries, without imposing explicit constraints on the network's architecture. In this work, for every spin configuration in the training set, we apply a global rotation or translation to get a new spin configuration with the same free energy. Since the rotation invariance is due to a continuous symmetry, in practice we had to choose a discrete set of Euler angles to implement the rotation-based data augmentation. More specifically, we use the twenty-three combinations of Euler angles  $\alpha = \{0, \pi\}$ ,  $\beta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ , and  $\gamma = \{0, \pi\}$ , where the  $\{\alpha = 0, \beta = 0, \gamma = 0\}$  combination is excluded since it corresponds to the initial spin configuration. Another approach to utilize a system's symmetries is to construct an equivariant neural network [36,37]. An equivariant neural network guarantees that the model obeys the system's symmetries using weight sharing and activation functions that respect those symmetries. An example of an equivariant neural network is a convolutional neural network which respects translation invariance. We opted for data augmentation due

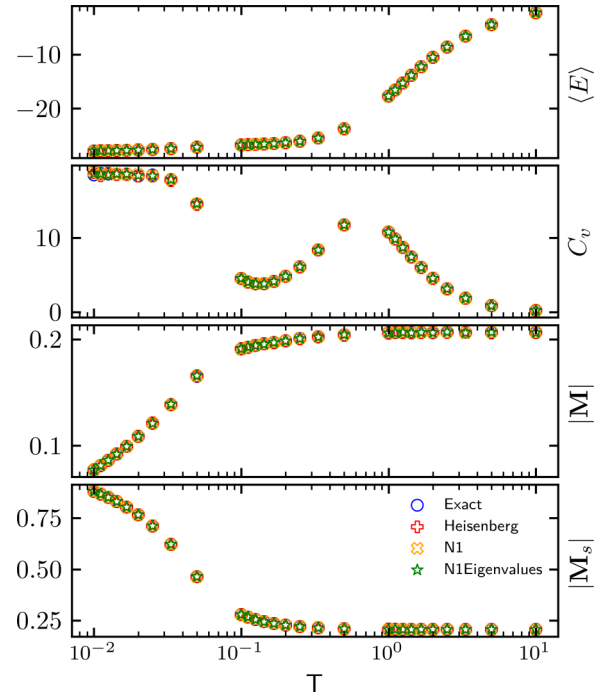


FIG. 2. Average energy, specific heat, magnitude of average magnetization, and staggered magnetization for an one-dimensional system with  $N = 20$  lattice sites. We generated  $10^5$  samples using exact diagonalization (blue circles), Heisenberg model (red pluses), N1 (orange crosses), and N1Eigenvalues (green stars). All models are in excellent agreement with the exact results.

to its much simpler and straightforward implementation. Another advantage of data augmentation is the ability to use the same neural network architecture for Hamiltonians that do not share the same symmetries.

At each temperature, we train two different neural networks that share the same architecture, as they are both fully connected feedforward networks with a single hidden layer, and measure their prediction accuracy using mean squared error on the test data set. The first network, which we refer to as N1, takes as input the classical spin components and outputs the free energy at that specific temperature. The second neural network, which we refer to as N1Eigenvalues, also takes as input the classical spin components, but outputs the Hamiltonian's eigenvalues. For the N1 model, we compare the predicted free energy and the actual free energy, whereas for the NEigenvalues model, we compare the predicted eigenvalues against the actual eigenvalues in the loss function, which is the MAE for both models. For both neural networks, we choose the architecture that combines the smallest mean squared error on a validation set with the least amount of hidden nodes in order to avoid overfitting and allow for a fast sample generation. For the one-dimensional system, we use  $h = 60$  hidden nodes and  $n_b = 20$  minibatches for N1 and  $h = 80$  hidden nodes and  $n_b = 50$  minibatches for N1Eigenvalues. For the two-dimensional system, we use  $h = 108$  hidden nodes and  $n_b = 20$  minibatches for N1 and  $h = 144$  hidden nodes and  $n_b = 20$  minibatches for N1Eigenvalues. In all cases, we observe that the mean squared error on the test data set decreases by starting at high

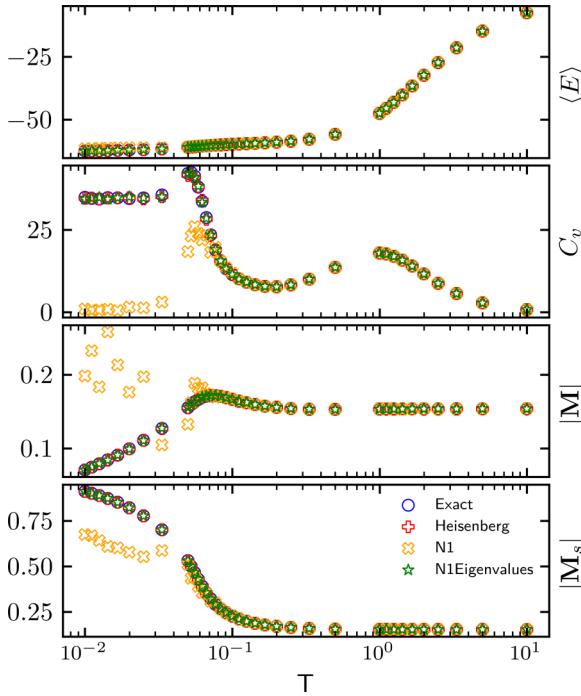


FIG. 3. Average energy, specific heat, magnitude of average magnetization, and staggered magnetization for a two-dimensional system with  $N = 6 \times 6 = 36$  lattice sites. We generated  $10^5$  samples using exact diagonalization (blue circles), Heisenberg model (red pluses), N1 (orange crosses), and N1Eigenvalues (green stars). The N1 model fails to properly describe the system's behavior in the low-temperature regime.

temperatures and moving sequentially to lower temperatures, using the optimized parameters corresponding to the previous temperature's model as the initial parameters for the next temperature's model.

For the one-dimensional system with  $N = 20$  sites, all models have a low mean squared error for the entire range of temperatures, with the linear Heisenberg model having the smallest over the entire temperature range (see Fig. 1). For the two-dimensional system with  $N = 6 \times 6 = 36$  sites, all models again have a low mean squared error. The Heisenberg model performs the best at high temperatures whereas the neural networks perform better at lower temperatures (see Fig. 1). For both systems, the maximum mean squared error reaches  $\sim 10^{-6}$  around  $T = 0.05$  where the average fermionic free energy is  $\mathcal{O}(1)$  indicating that our models are able to predict the fermionic free energy with high accuracy.

#### IV. IMPORTANCE SAMPLING

The integrals in Eqs. (2), (5), (6), (7), and (8) cannot be evaluated in a closed form, requiring us to resort in numerical methods. If the probability distribution from Eq. (3) is known, then the expected value  $\langle O \rangle$  of a function  $O(\mathbf{S})$  is

$$\langle O(\beta) \rangle = \int d\mathbf{S} O(\mathbf{S}) p(\beta, \mathbf{S}), \quad (14)$$

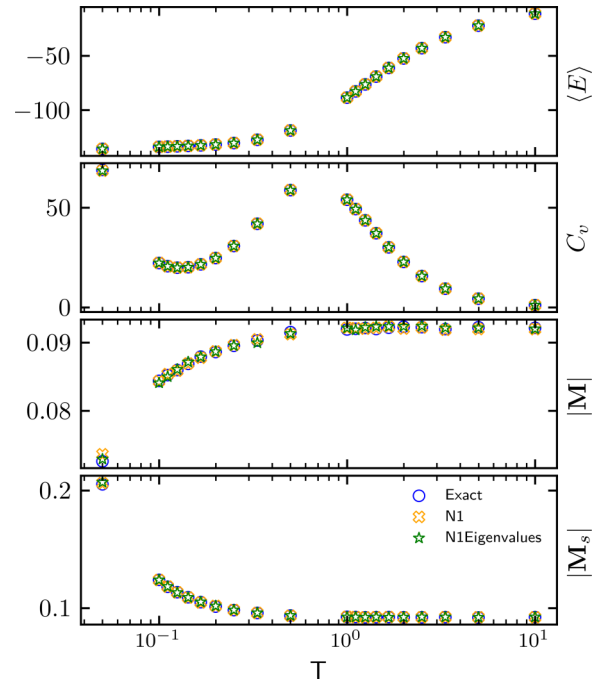


FIG. 4. Average energy, specific heat, magnitude of average magnetization, and staggered magnetization for an one-dimensional system with  $N = 100$  lattice sites. We generated  $10^5$  samples using exact diagonalization (blue circles), N1 (orange crosses), and N1Eigenvalues (green stars). Similarly with its smaller counterpart ( $N = 20$ ), all models are in excellent agreement with the exact results.

and can be approximated with

$$\langle O(\beta) \rangle \approx \frac{1}{N} \sum_{i=1}^N O(\mathbf{S}_i), \quad (15)$$

using spin configurations generated according to Eq. (3). However, it is computationally expensive to generate samples according to the exact probability distribution governing our system, which is why we resort to using an effective model. More specifically, given the probability distribution according to the effective model  $q(\beta, \mathbf{S})$ , the expected value  $\langle O(\beta) \rangle$  is

$$\langle O(\beta) \rangle = \int d\mathbf{S} O(\mathbf{S}) q(\mathbf{S}) \frac{p(\mathbf{S})}{q(\mathbf{S})}, \quad (16)$$

and can be approximated with

$$\langle O(\beta) \rangle \approx \frac{\sum_{i=1}^n w(\beta, \mathbf{S}_i) O(\mathbf{S}_i)}{\sum_{i=1}^n w(\beta, \mathbf{S}_i)}, \quad (17)$$

where we used

$$w(\beta, \mathbf{S}) = e^{-\beta(\mathcal{F}(\beta, \mathbf{S}) - \hat{\mathcal{F}}(\beta, \mathbf{S}))}, \quad (18)$$

with  $\mathcal{F}(\beta, \mathbf{S})$  and  $\hat{\mathcal{F}}(\beta, \mathbf{S})$  being the fermionic free energy due to the exact and effective models, respectively. This process is referred to as *importance sampling* in literature [38]. It is worthwhile noting that the estimator in Eq. (17) is asymptotically unbiased [39]. Calculating Eq. (18) requires us to diagonalize the Hamiltonian for a given spin configuration. One might reasonably ask what computational gains do we get by using an effective model if we still need

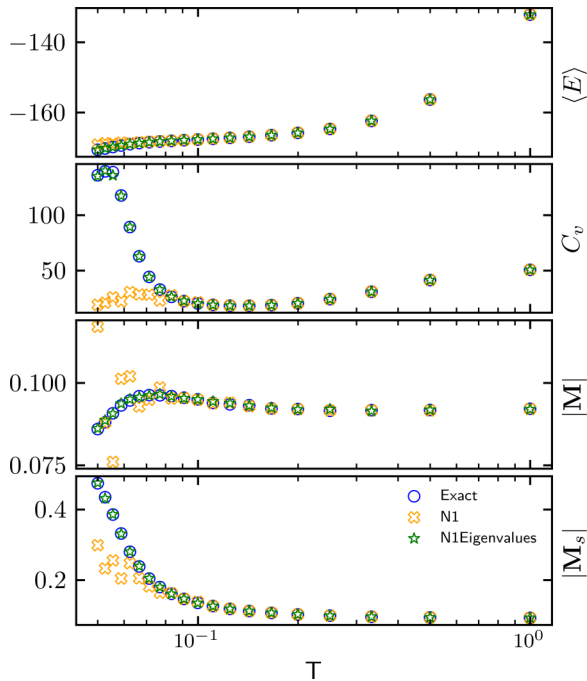


FIG. 5. Average energy, specific heat, magnitude of average magnetization, and staggered magnetization for a two-dimensional system with  $N = 10 \times 10 = 100$  lattice sites. We generated  $10^5$  samples using exact diagonalization (blue circles), N1 (orange crosses), and N1Eigenvalues (green stars). Similarly with the smaller counterpart of  $N = 6 \times 6 = 36$  the N1 model fails to properly describe the system's behavior in the low-temperature regime.

to perform exact diagonalization for importance sampling? The computational savings come during the decorrelation stage of the METROPOLIS Hastings algorithm where for every  $N/r$  spin configurations generated we record only the last one so using an effective model for all those unrecorded update steps reduces the time needed for sample generation.

For the one-dimensional system ( $N = 20$  lattice sites), the effective models generate samples that accurately describe the average energy, specific heat, the magnitude of average magnetization, and the staggered magnetization as can be seen in Fig. 2. For the Heisenberg model, we do not generate samples below  $T = 10^{-2}$  because the sample generation script requires approximately the same amount of time to generate samples as exact diagonalization, a fact we attribute to our code not being properly optimized.

The situation is different for the two-dimensional case ( $N = 6 \times 6 = 36$  lattice sites). All models generate samples that are able to accurately describe the average energy, but the N1 model fails to capture the correct behavior for the other three quantities in the region of  $T \lesssim 0.1$  as can be seen in Fig. 3. This seems counterintuitive given the extremely low mean squared error that N1 has as can be seen in Fig. 1, but this is an indication of overfitting. We have tried increasing both the number of minibatches used during training and the number of hidden nodes for the N1 model; however the results remained practically the same. The figures for the spin correlation and structure factor for the systems under consideration are presented in Appendix C.

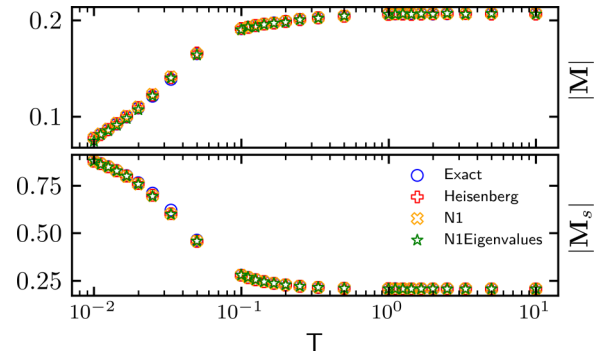


FIG. 6. Magnitude of average magnetization and staggered magnetization for the one-dimensional system ( $N = 20$  lattice sites) without implementing importance sampling. All three models maintain the same performance as shown in Fig. 2.

We are also interested in testing the viability of our method on larger systems. We focus our attention on a one-dimensional chain with  $N = 100$  sites and a two-dimensional square lattice with  $N = 10 \times 10 = 100$  sites. Both systems are significantly larger than the ones studied in Sec. IV, while at the same time allow us to generate enough training data in a reasonable amount of time. We use the same approach as the one outlined in Sec. III B, with the only difference being the number of hidden nodes used in each neural network. For each model, we use the same ratio of hidden nodes to inputs as in the previous cases. More specifically, we use 1 : 1 and 4 : 3 hidden nodes to inputs for the N1 and N1Eigenvalues respectively. As can be seen in Figs. 4 and 5, the neural networks maintain the same performance as that demonstrated in their smaller counterparts. We should note that we use a training data set with  $10^4$  samples for both large and small lattices and believe it is a serendipitous fact that the models perform in a similar manner. We expect that training these model architectures on larger systems with  $\sim 1000$  sites will probably require bigger training data set to achieve similar performance.

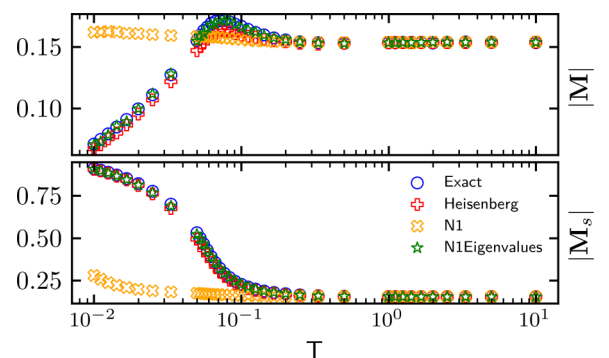


FIG. 7. Magnitude of average magnetization and staggered magnetization for the two-dimensional system ( $N = 6 \times 6 = 36$  lattice sites) without implementing importance sampling. The N1Eigenvalues maintains the same performance as shown in Fig. 3, whereas the Heisenberg model shows a slight disagreement in the region around  $T \sim 0.1$ . The N1 model completely fails for temperatures below  $T \sim 0.2$ .

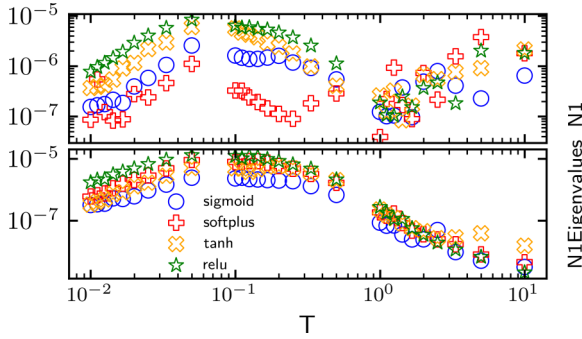


FIG. 8. Comparing the MSE using different activation functions for the N1 and N1Eigenvalues models. We compare the sigmoid function (blue circles), the softplus function (red pluses), the hyperbolic tangent function (orange crosses), and the relu function (green stars).

The main bottleneck of the methods presented in this paper is the amount of time needed to generate the training data set since this is obtained using exact diagonalization. Thus, one could reasonably question if the amount of time needed to train a neural network and use it to generate samples is worth the effort. We investigated this question and found that generating a training data set, training a neural network, and subsequently using it to generate samples requires less time than generating samples using exact diagonalization. For this benchmark, we devoted the same resources to both approaches; however, one has to keep in mind that the absolute time will differ on a different machine. For the one-dimensional system with  $N = 100$  sites, it takes

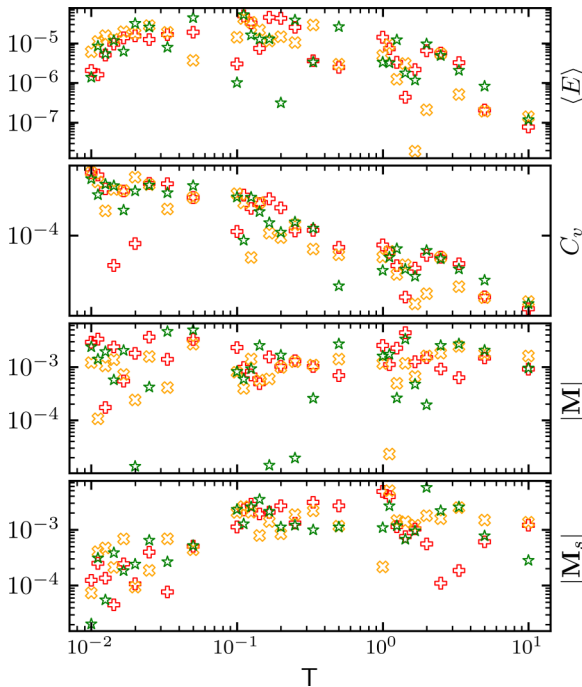


FIG. 9. The relative error between the Heisenberg model (red pluses), N1 (orange crosses), N1Eigenvalues (green stars), and the quantities generated with exact diagonalization for the one-dimensional system with  $N = 20$  sites.

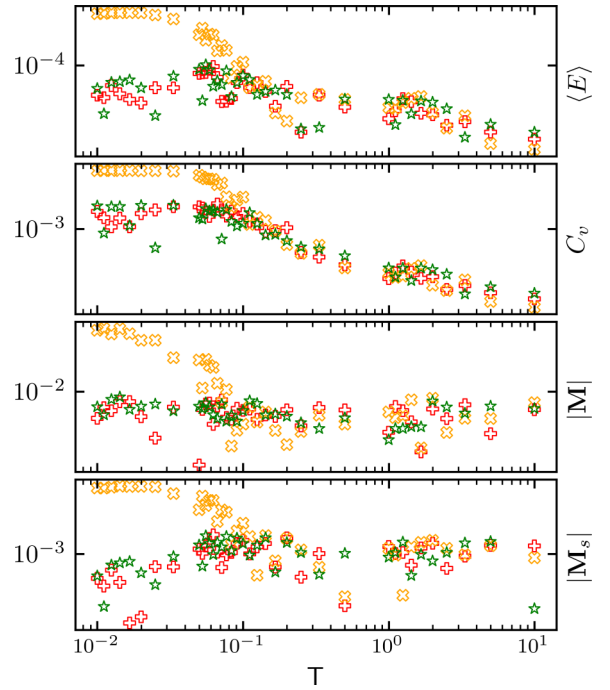


FIG. 10. The relative error between the Heisenberg model (red pluses), N1 (orange crosses), N1Eigenvalues (green stars), and the quantities generated with exact diagonalization for the two-dimensional system with  $N = 6 \times 6 = 36$  sites.

approximately 4.5 hours to generate the training data set with  $10^4$  samples at  $T = 0.1$ . Both neural networks are trained using a GPU, which expedites the training process, and that takes about 90 minutes regardless of the temperature. The neural networks require approximately 25 minutes to generate  $10^5$  samples at  $T = 0.1$ . Hence the total amount needed to generate  $10^5$  samples at  $T = 0.1$  using the neural networks, including the amount of time needed to generate the training data set and train the models, is roughly 6.5 hours. In contrast, to generate  $10^5$  samples at  $T = 0.1$  using exact diagonalization requires roughly 41 hours. It is worth noting that the amount of time required to train the neural network depends on the number of epochs, batch size, and training data set size. For instance, one could reduce the training time by using fewer epochs, a bigger batch size, or a smaller training data set. In the future, we would like to investigate the possibility of training a neural network on a smaller lattice and appropriately adjust it so it can generate samples for a larger lattice.

## V. NEURAL NETWORKS AS EFFECTIVE MODELS

Lastly, we examine the possibility of completely eliminating exact diagonalization from the calculation of the expected values. That is equivalent to using Eq. (15) where the spin configurations are drawn from the probability distribution derived from the effective models. In other words, neural networks can be seen as *nonperturbative* effective models obtained by tracing out over the electronic degrees of freedom.

To test this idea, we calculate the observables related to the magnetization state of the system since both the average energy and specific heat require knowledge of the

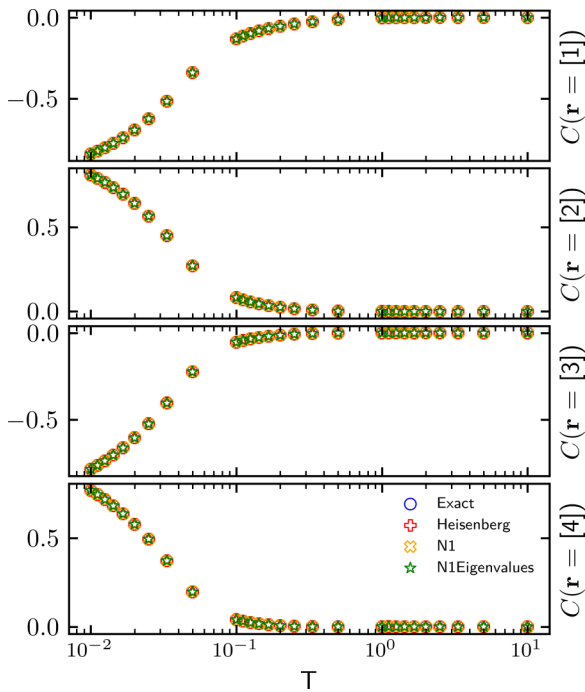


FIG. 11. Spin correlation for an one-dimensional system with  $N = 20$  lattice sites for  $\mathbf{r} = \{1, 2, 3, 4\}$ .

eigenspectrum which cannot be accessed with the N1 and Heisenberg models. Having said that, we could still use N1Eigenvalues for determining these two quantities since its output is the eigenspectrum of the Hamiltonian. For the one-dimensional system (see Fig. 6), all three models remain in agreement with exact diagonalization maintaining the same performance as in Fig. 2. For the two-dimensional system (see Fig. 7), only N1Eigenvalues remains in agreement with exact diagonalization across all temperatures. Interestingly, the Heisenberg model remains in agreement with exact diagonalization for most temperatures, and struggles only in the region around  $T \sim 0.1$ . The N1 completely fails to capture the correct behavior for temperatures below  $T \sim 0.2$ , which is precisely the regime where the physics is dominated by the energy spacing. On the other hand, it is reasonable to expect that N1Eigenvalues will be better to retain its high fidelity even in the absence of importance sampling since the network approximates the mapping of each spin configuration to its corresponding eigenspectrum, a task which might preserve a significant amount of information relevant for describing the system. In contrast, both the Heisenberg and N1 models try to approximate the fermionic free energy, a scalar quantity obtained by integrating over the spectrum, a process that might cause significant information loss.

## VI. CONCLUSION

Neural networks are ideal surrogates to replace computationally intensive steps in Monte Carlo methods, given their remarkable flexibility to approximate arbitrary functions of interest. This advantage is counterbalanced by the numerical cost required to train the models. One solution is to exploit a system's symmetries and augment the existing training data

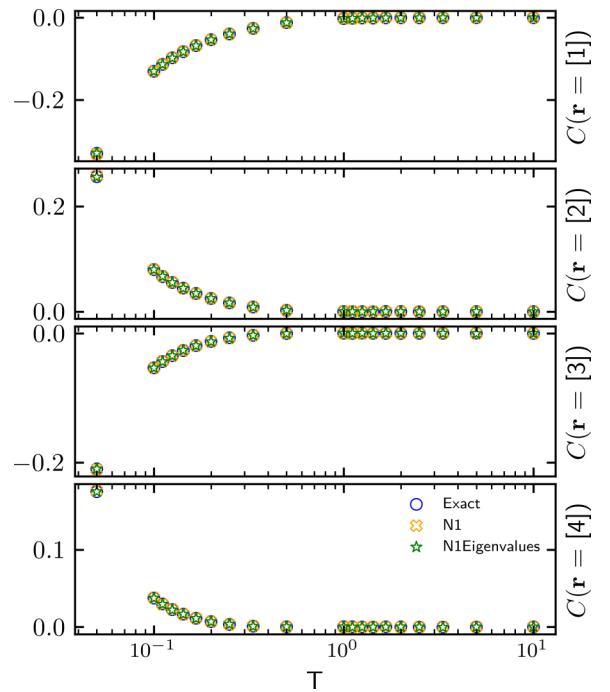


FIG. 12. Spin correlation for an one-dimensional system with  $N = 100$  lattice sites for  $\mathbf{r} = \{1, 2, 3, 4\}$ .

set, thus creating a bigger data set without the overhead of the exact method. An added benefit of data augmentation is that we are actively encouraging the neural network to learn the system's symmetries that the model should also obey. In this work, we train two neural network models using data augmentation and demonstrate their ability to predict the free energy for the spin-fermion model in one and two dimensions. The two neural network models differ fundamentally in two aspects: (i) the quantity they learn and (ii) how they are used as effective models to generate samples. The N1 model learns the free energy and is used in combination with importance sampling which requires to diagonalize the Hamiltonian at each measurement step. The N1Eigenvalues model learns the energy spectrum (or density of states) of the Hamiltonian, and can reproduce the exact results with high level of accuracy without using importance sampling, making the simulations considerably more efficient and faster. We compare the two neural networks against an effective Heisenberg model and find that in one dimension all three models have comparable performance, whereas in two dimensions only the N1Eigenvalues and the Heisenberg model describe correctly the system's behavior across all temperatures. Unlike the Heisenberg model, the N1Eigenvalues model is not constrained to have a specific functional form and is able to find an appropriate approximation to the energy spectrum. Lastly, since the energy spectrum is temperature independent, a neural network that learns the energy spectrum, such as the N1Eigenvalues model, could be trained at a high temperature where it is cheap to generate a large training data set and then used for all temperatures. Such an approach will reduce the cost associated with training allowing the study of larger systems at low temperatures.



**ACKNOWLEDGMENTS**

We acknowledge the support we received from NU TIER1 FY21 and the National Science Foundation under Awards ECCS-1845833 (T.I. and P.C.) and DMR-2120501 (A.E.F.). G.S. was supported by the Roux Institute at Northeastern University and the Harold Alfond Foundation.

**APPENDIX A: COMPARING ACTIVATION FUNCTIONS**

While constructing both neural networks we investigated how much our results were affected by the choice of the activation function. We compared the following functions: (a) the sigmoid function  $f(x) = \frac{1}{1+e^{-x}}$ , (b) the softplus function  $f(x) = \ln(1 + e^x)$ , (c) the hyperbolic tangent function  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , and (d) relu function  $f(x) = \max(0, x)$ . Our results are shown in Fig. 8, where we used the same number of hidden nodes and training epochs. For the N1 network, the softplus seems to perform the best over most temperatures, whereas for the N1Eigenvalues the sigmoid has the best performance. However, all activation functions have extremely low MSE over all temperatures. We choose to use softplus for both N1 and N1Eigenvalues so they are on equal footing when we compare them.

**APPENDIX B: MODEL PERFORMANCE**

Even though the results presented in Figs. 2 and 3 are in good agreement with those generated using exact diagonalization, it is hard to see by how much the quantities as determined using the different models deviate from those derived using

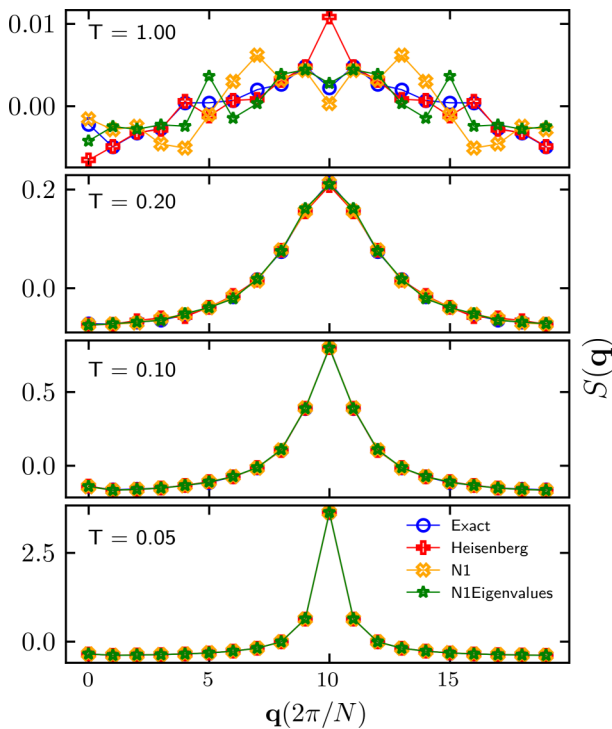


FIG. 13. Spin structure factor for an one-dimensional system with  $N = 20$  lattice sites for  $T = \{1, 0.2, 0.1, 0.05\}$ . We observe a peak at  $\mathbf{q} = 10 \times \frac{2\pi}{20} = \pi$  at  $T = 0.05$ . All models are in good agreement with exact diagonalization.

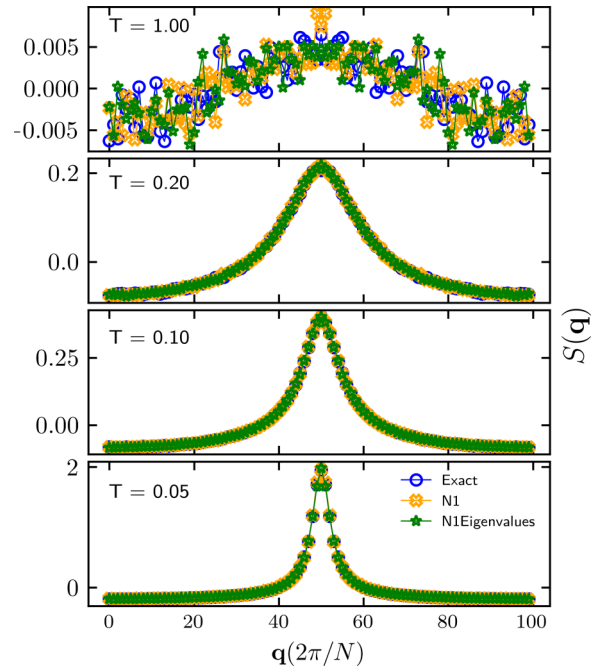


FIG. 14. Spin structure factor for an one-dimensional system with  $N = 100$  lattice sites for  $T = \{1, 0.2, 0.1, 0.05\}$ . We observe a peak at  $\mathbf{q} = 50 \times \frac{2\pi}{100} = \pi$  at  $T = 0.05$ . All models are in agreement with exact diagonalization.

exact diagonalization. In order to further probe the accuracy of our models, we generated Figs. 9 and 10 that present the relative error for the four quantities of interest, for two of the systems under consideration. In the one-dimensional case, all models have a really low relative error, but in the two-dimensional system the N1 model has a high relative error at low temperatures.

**APPENDIX C: SPIN CORRELATION AND STRUCTURE FACTOR**

The spin correlation for the one-dimensional  $N = 20$  and  $N = 100$  systems are shown in Figs. 11 and 12 respectively for  $\mathbf{r} = \{1, 2, 3, 4\}$ . In both systems, we notice that the magnitude of the correlation increases as the temperature approaches zero. Furthermore, the spin correlation associated with the odd vectors is negative whereas for the even vectors is positive. This is an indication of antiferromagnetism and is confirmed by the fact that the staggered magnetization approaches unity and the average magnetization approaches zero as the temperature approaches zero.

For these two systems, we also calculate the spin structure factor and the results are shown in Fig. 13 and 14 for  $\mathbf{q} = \frac{2\pi}{N} \{0, 1, \dots, L - 1\}$ . One can see that at high temperature ( $T = 1$ ) the structure factor seems relatively flat and the models struggle to properly describe it. However, at a low temperature ( $T = 0.05$ ), we see a pronounced peak at  $\mathbf{q} = \pi$  and all models are in good agreement with exact diagonalization.

The spin correlation for the two-dimensional  $N = 6 \times 6 = 36$  and  $N = 10 \times 10 = 100$  square lattices are shown in Figs. 15 and 16 respectively for  $\mathbf{r} = \{(0, 1), (1, 0), (1, 1), (0, 2)\}$ . Similarly with the one-dimensional case the magnitude for all spin correlations

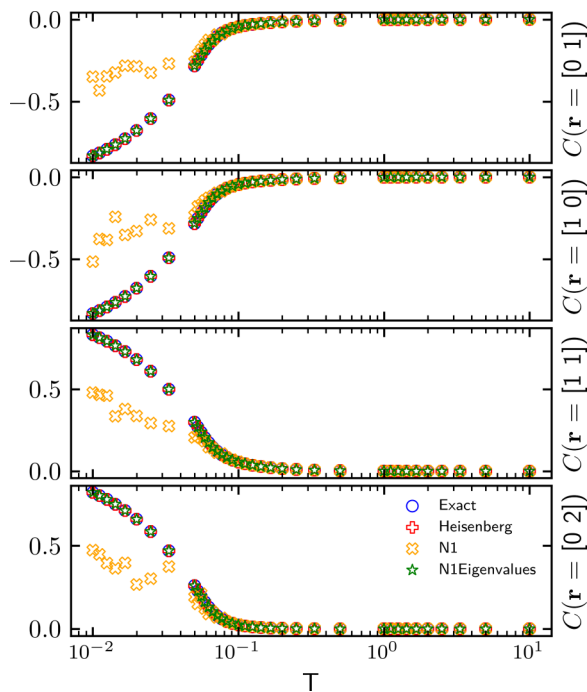


FIG. 15. Spin correlation for a two-dimensional system with  $N = 6 \times 6 = 36$  lattice sites for  $\mathbf{r} = \{(0, 1), (1, 0), (1, 1), (0, 2)\}$ .

increases as  $T \rightarrow 0$  and the systems acquire a checkerboard pattern thus exhibiting antiferromagnetic ordering.

For these two systems, we also calculate the spin structure factor and the results are shown in Figs. 17 and 18. Similarly to the one-dimensional case, we observe that at high temperature ( $T = 1$ ) the structure factor is relatively flat and the models struggle to properly describe it. At a low temperature

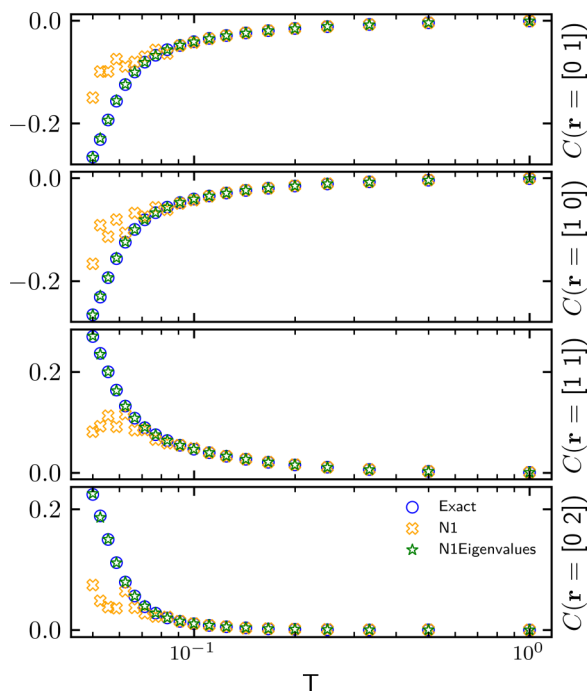


FIG. 16. Spin correlation for a two-dimensional system with  $N = 10 \times 10 = 100$  lattice sites for  $\mathbf{r} = \{(0, 1), (1, 0), (1, 1), (0, 2)\}$ .

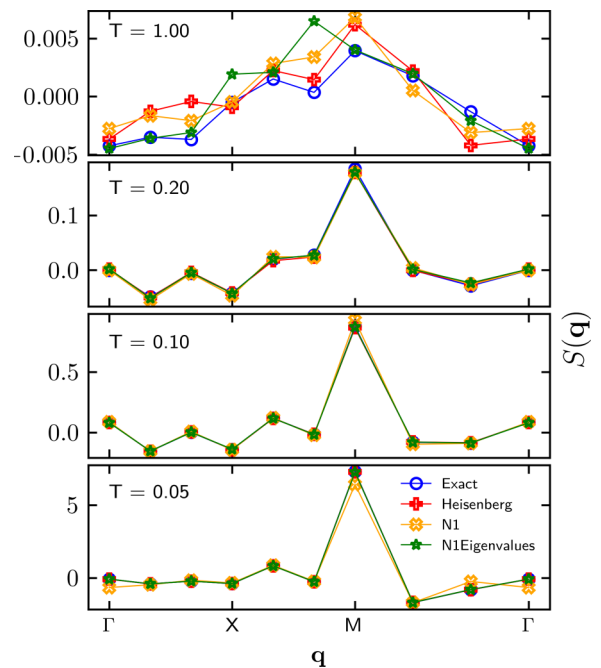


FIG. 17. Spin structure factor for a two-dimensional system with  $N = 6 \times 6 = 36$  lattice sites for  $T = \{1, 0.2, 0.1, 0.05\}$ . We observe a peak at  $\mathbf{q} = (\pi, \pi)$  (M) at  $T = 0.05$ . The Heisenberg and N1Eigenvalues models are in good agreement with exact diagonalization.

( $T = 0.05$ ), we see a pronounced peak at  $\mathbf{q} = (\pi, \pi)$  (M) with the Heisenberg and the N1Eigenvalues being in good agreement with exact diagonalization.

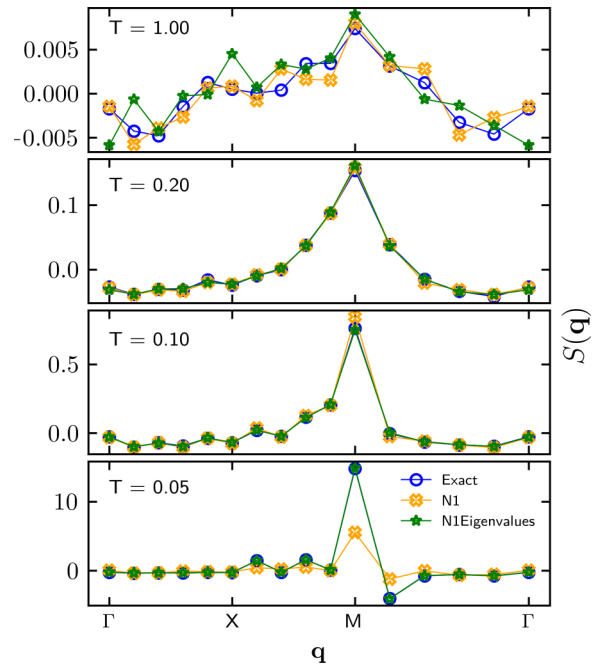


FIG. 18. Spin structure factor for a two-dimensional system with  $N = 10 \times 10 = 100$  lattice sites for  $T = \{1, 0.2, 0.1, 0.05\}$ . We observe a peak at  $\mathbf{q} = (\pi, \pi)$  (M) at  $T = 0.05$ . The N1Eigenvalues is in good agreement with exact diagonalization.

- [1] E. Morosan, D. Natelson, A. H. Nevidomskyy, and Q. Si, *Adv. Mater.* **24**, 4896 (2012).
- [2] *Strongly Correlated Systems*, edited by A. Avella and F. Mancini, Springer Series in Solid-State Sciences Vol. 176 (Springer, Berlin, Heidelberg, 2013).
- [3] *Strongly Correlated Systems: Experimental Techniques*, edited by A. Avella and F. Mancini, Springer Series in Solid-State Sciences Vol. 180 (Springer, Berlin, Heidelberg, 2015).
- [4] D. Ceperley and B. Alder, *Science* **231**, 555 (1986).
- [5] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems*, 1st ed. (Cambridge University Press, 2017).
- [6] C. Zener, *Phys. Rev.* **82**, 403 (1951).
- [7] P. W. Anderson and H. Hasegawa, *Phys. Rev.* **100**, 675 (1955).
- [8] P. G. de Gennes, *Phys. Rev.* **118**, 141 (1960).
- [9] K. Kubo and N. Ohata, *J. Phys. Soc. Jpn.* **33**, 21 (1972).
- [10] E. Dagotto, T. Hotta, and A. Moreo, *Phys. Rep.* **344**, 1 (2001).
- [11] T. Hotta and E. Dagotto, *Colossal Magnetoresistive Manganites*, edited by T. Chatterji (Kluwer Academic Publishers, Dordrecht, Netherlands, 2004), pp. 207–262.
- [12] E. Dagotto, *Nanoscale Phase Separation and Colossal Magnetoresistance: The Physics of Manganites and Related Compounds*, Springer Series in Solid-State Sciences Vol. 136 (Springer, Berlin, New York, 2003).
- [13] E. Berg, M. A. Metlitski, and S. Sachdev, *Science* **338**, 1606 (2012).
- [14] N. Furukawa, Y. Motome, and H. Nakata, *Comput. Phys. Commun.* **142**, 410 (2001).
- [15] N. Furukawa and Y. Motome, *J. Phys. Soc. Jpn.* **73**, 1482 (2004).
- [16] G. Alvarez, C. Şen, N. Furukawa, Y. Motome, and E. Dagotto, *Comput. Phys. Commun.* **168**, 32 (2005).
- [17] A. Weiße, *Phys. Rev. Lett.* **102**, 150604 (2009).
- [18] G. Alvarez, P. K. V. V. Nukala, and E. D’Azevedo, *J. Stat. Mech.* (2007) P08007.
- [19] K. Barros and Y. Kato, *Phys. Rev. B* **88**, 235101 (2013).
- [20] G. Torlai and R. G. Melko, *Phys. Rev. B* **94**, 165134 (2016).
- [21] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [22] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
- [23] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [24] D. Hendry and A. E. Feiguin, *Phys. Rev. B* **100**, 245123 (2019).
- [25] D. Hendry, H. Chen, P. Weinberg, and A. E. Feiguin, *Phys. Rev. B* **104**, 205130 (2021).
- [26] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, *Phys. Rev. B* **95**, 041101(R) (2017).
- [27] L. Huang and L. Wang, *Phys. Rev. B* **95**, 035105 (2017).
- [28] H. Shen, J. Liu, and L. Fu, *Phys. Rev. B* **97**, 205140 (2018).
- [29] D. Wu, R. Rossi, and G. Carleo, *Phys. Rev. Res.* **3**, L042024 (2021).
- [30] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [31] M. A. Ruderman and C. Kittel, *Phys. Rev.* **96**, 99 (1954).
- [32] K. Yosida, *Phys. Rev.* **106**, 893 (1957).
- [33] H. Kohshiro and Y. Nagai, *J. Phys. Soc. Jpn.* **90**, 034711 (2021).
- [34] D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [35] T. Dao, A. Gu, A. J. Ratner, V. Smith, C. De Sa, and C. Ré, [arXiv:1803.06084](https://arxiv.org/abs/1803.06084).
- [36] T. S. Cohen and M. Welling, [arXiv:1602.07576](https://arxiv.org/abs/1602.07576).
- [37] R. Wang, R. Walters, and R. Yu, [arXiv:2002.03061](https://arxiv.org/abs/2002.03061).
- [38] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics (Springer, New York, NY, 2004).
- [39] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez, and P. M. Djuric, *IEEE Signal Process. Mag.* **34**, 60 (2017).